

H04Q11/0452
T04Q11:0452B22
T04Q11:0452B16
T04Q11:0452B20

807

TELETRAFFIC AND DATATRAFFIC
in a Period of Change, ITC-13
A. Jensen and V.B. Iversen (Editors)
Elsevier Science Publishers B.V. (North-Holland)
© IAC, 1991



XP 000303113

ANALYTICAL METHODS FOR THE TRAFFICAL PROBLEMS WITH STATISTICAL MULTIPLEXING IN ATM-NETWORKS

Karl LINDBERGER

SWEDISH TELECOM NETWORKS, S-123 86 FARSTA SWEDEN

p.807 - 813

This paper is discussing statistical multiplexing on the links in an ATM-network. We are giving the series of different steps of the description of the traffic processes and their effect on the cell delay and the cell loss probability. The results are then meant to be used for call acceptance and charging principles.

1. INTRODUCTION

Traffical problems with statistical multiplexing will be treated in this paper. Most of the ideas here are originally from some of my internal papers [1]-[5] in COST224 or from discussions there during the last three years. Now some of the suggested methods might have their equivalences in papers by other authors while others could still have great originality when finally presented in ITC. Nevertheless, to show the structure, the whole series of steps in the mathematical treatment of the statistically multiplexed traffic on the ATM-links will be given here with my own variants of methods as examples.

Even though we in reality have a traffic process on the link going on only in one discrete time scale it has appeared to be practical to describe this process in three different time scales, the cell- the burst- and the call-scales, since this traffic process has different types of variations for different reasons. In section 2 we shall comment on the so called cell scale problems which are determining the load of the link and the buffer size for a certain GOS also in, or even particularly in, cases without statistical multiplexing.

In section 3 we shall suggest different models to describe the offered traffic processes with their variation in the burst scale which we consider being the most important one for the study of statistical multiplexing. We shall also study cell loss probabilities given a certain link capacity. In section 4 we consider a system with a great number of buffers which gives a problem with significant delay also in this burst scale sense. We can then decide whether a greater number of buffers are worthwhile or not. The results from section 3 can then be summarized in a way where the individual resource allocation for a certain type of call is approximately independent of the mixture of calls we have on the link. This almost linear quality holds rather well in the part of

the parameter space where statistical multiplexing is interesting in the first place. How this linearity can be expressed is shown in section 5. With this virtual resource allocation value for each call we have a call acceptance situation in the call scale which is similar to the one in circuit switched multi-slot networks. Then in section 6 we shall find some possible combined principles for control, policing, measurements and charging, where we can handle an offered call with a certain peak rate and a declared average rate. If that is the only thing we know, the call acceptance principle is not obvious. Can we trust the declared average rate? How can it be checked? Finally in section 7 we shall summarize which savings there are in statistical multiplexing.

2. THE CELL SCALE PROBLEMS

The cell scale problems shall not be treated in detail in this paper. However we shall comment on their formulation and their major results and later how those should be combined with the burst scale results. These problems are important also, or even- particularly in ATM without statistical multiplexing. They are checking the very local (in time) variation caused by the fact that when a lot of small independent streams are offered to the same link there is a certain chance that too many of those might arrive at almost the same time slot or adjacent time slots instead of being uniformly spread out. This problem is solved by a reasonable number of buffers and by not loading the link too close to 100 %. The importance of this effect is dominating when we have many small streams of burstiness factors close to 1. For cases with bursty streams from a smaller number of sources this additional local variance will however be dominated by the 'burst scale variance' and then we will not be able to have a load of the link close to 100 % anyway. However the sort of problem we study here is of the types $nD|D|1/k$ or $\Sigma D_i|D|1/k$. The first one can be solved exactly and for the other one we

TELETRAFFIC AND DATATRAFFIC

in a Period of Change

ITC-13

Proceedings of the Thirteenth International Teletraffic Congress
Copenhagen, Denmark, June 19-26, 1991

Edited by

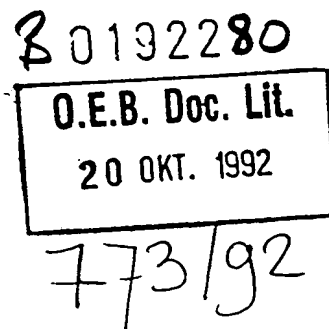
Arne JENSEN and V. B. IVERSEN

Technical University of Denmark
Lyngby, Denmark



1991

NORTH-HOLLAND — AMSTERDAM • LONDON • NEW YORK • TOKYO



refer to approximations, see e.g. [10] and [11]. The type of answers we get is: given e.g. a GOS B-10⁻⁹ and the buffer size k=100, we can load the link to e.g. 90 % without statistical multiplexing (i.e. burstiness factor 1) for a certain type of sources. Several authors have concentrated on these problems.

3. THE BURST SCALE PROBLEM

In cases with statistical multiplexing it is rather obvious that the variations we can observe in the so called 'burst scale' are the interesting ones. We choose a time unit which is essentially greater than a time slot for a cell e.g. 1 ms (in the right magnitude though other choices could be suggested) and do not here worry about any possible variations within those ms:s. The situation we shall study is one where we have conditioned a certain number of calls of each type going on. The variation in number of calls of each type is later discussed in the call scale. A general traffic stream, either it is one with variable bit rate or a pure burst process is characterized by a certain number of moments. A suggestion with e.g. three moments and 1 ms as the time unit could be the following (see [1]).

m, σ^2 : the mean and the variance of the number of offered cells per ms and

v : where vt is the asymptotic variance of the number of offered cells in a long time interval of length t .

The advantage with this choice of parameters is that the corresponding ones for the superposition process of several independent streams are obtained as the sum of the individual ones, e.g. for two streams:

$$m = m_1 + m_2 \quad \sigma^2 = \sigma_1^2 + \sigma_2^2 \quad v = v_1 + v_2$$

The general offered traffic stream is then replaced in the dimensioning problem by an equivalent process which is a simple pure burst process with the same parameter values, i.e. a special case in the class of general processes.

The equivalent processes are given in two versions, one based on the assumption of infinitely many sources and the other one having a finite number of sources and thus needing an extra parameter to fit.

I The first traffic process has a Poissonian arrival process (λ) for independent but equally distributed bursts with fixed height h cell/ms and burst durations exponentially distributed with mean d . Thus the average burst size is $b = dh$. The process is determined by (m, h, b) and the fitting relations are

$$h = \sigma^2/m, \quad b = v/2m$$

The relation $\lambda b = m$ can also be noted. The expression for v follows from the application of the formula for the asymptotic variance vt of a cumulative regenerative process, see [8] or [7] p.164 random sums.

II The second type of equivalent offered traffic process has a limited number, M , of equally distributed independent sources each one with exponentially distributed on- and off-times for bursts with fixed height h and mean burst duration d . The relation between on- and off-times for a source can be expressed by the parameter mean proportion of on-time a . The fitting relations based on the three moments are now (see [8] again).

$$m = Mah, \quad \sigma^2 = Ma(1-a)h^2, \quad v = 2Ma(1-a)^2dh^2$$

Thus we have 3 relations but 4 parameters to fit. In [1] is suggested an additional relation

$$ma = \sum m_i a_i$$

where a_i is the individual traffic for source i which we would have got in a case with a known number M_i of only this type of sources. In [12] on the other hand the same 4 parameters are also adopted only with the alternative fourth fitting relation $M = \sum M_i$ i.e. the same number of sources in the equivalent stream as in the offered one.

As a variant to this model with discrete states for the equivalent bursts is also suggested in [15] a model with densities varying continuously as an Ornstein-Uhlenbeck process with similar moments as the key ones.

Now we have the offered models I with the parameters (m, h, b) and II with (M, a, h, d) where $m = Mah, b = dh$. In [1] we found that when $R/b \ll 1$ i.e. the number of buffers was small in relation to the standard burst size, then the blocking problem was hardly including any essential waiting possibilities for bursts and thus the exact values of R and b were unimportant. The cell blocking probabilities could then be expressed as

$$B = E(X-n | X \geq n) P(X \geq n) / E(X)$$

where $X \sim \text{Po}(m/h)$ in I, $X \sim \text{Bi}(M, a)$ in II

and $n = c/h$ with c as the link capacity

The blocking quantity is based on the burst scale model of the problem with no buffers at all ($R = 0$), and the fact that a burst arriving when the link capacity is already fully used will not always loose all its cells (as many cells as in a burst) but often only some of the cells in the beginning of the burst. That is why B is not an ordinary Erlang or Engset formula.

Other variants of approximation are also possible e.g. letting the equivalent distribution i.e. the correspondence to X be either the sum of a constant and Poissonian random variable which takes 3 moments to fit or simply a normal distribution which takes only 2. Whatever method we choose of this simplicity there is however always possible to find certain cases e.g. mixtures of streams from a few big sources and many small sources, where the accuracy is not good enough at the level of e.g. $B = 10^{-9}$. Usually those cases are not very efficient for statistical multiplexing anyway i.e. the big sources might not have offered any significant savings on a certain link even when integrated only with each other. It could also be discussed whether $B = 10^{-9}$ should be regarded as a magnitude or as $1.0 \cdot 10^{-9}$.

Now the two expressions for B can of course be calculated numerically in I and II with either Poisson or Binomial distribution for X . In the case of I we can also suggest an analytical approximation for not too small n -values based on

$$P(X=n+k) = \rho^k P(X=n)$$

where

$$X \sim \text{Po}(n\rho), \quad n=c/h \quad \rho=m/c$$

Then it can be shown that

$$B \approx n^{-1}(1-\rho)^{-2} P(X=n)$$

This expression, though not exact, shows the essential difference between B and e.g. the Erlang formula which can not be used in this case. This method I is simple in the sense that once you have determined the general peak rate h of the bursts in the equivalent process for a certain mixture and thus also $n = c/h$ you can use methods above and construct a simple table e.g. for $B = 10^{-9}$.

n	20	30	50	60	75	100	150	300
ρ	21	29	41	45	50	55	62	72

For a certain n -value you can see which load of the link in percents you can offer for the particular mixture which gives the h -value. For small values of n however we suggest the method II.

So far we have assumed that all moments can be measured. In these early discussions however we don't have any measurements but only ask about dimensioning for mixtures of streams with certain peak and average rates. With only that information known about the streams, we must assume that they are simple on-off stream so that we can find a correspondence between the peak rate of a stream and its variance σ_i^2 . We could now use method II but if we still prefer to apply the simple method I we must at least modify the peak rate.

The real peak rate h is the right one for II i.e. $h_{II} = h$ but for h_I in I it holds

$$h_I = h(1-a) = h-m = \Sigma(h_i-m_i)$$

Otherwise it won't fit with the variances. With this h_I in $n_I = c/h_I$ it is more reasonable to use the table for I.

As an example we could consider mixtures of two types of calls with peak rates 11 and 3 Mbps respectively and both having average rates 1 Mbps i.e. they have very different burstiness factors. Now these real peak rates correspond to the model I burst peak rates $h_I = h_i-m_i$ i.e. $11-1 = 10$ and $3-1 = 2$ which we now can use in model I.

Some possible mixtures of these two services on a 600 Mbps-link are given in this table.

Prop. $h_I=10$, $h_{II}=11$ calls	n	ρ	Number of calls		
			10(11)	2(3)	Total
1.00	60	0.45	270	0	270
0.75	75	0.50	223	74	297
0.50	100	0.55	165	165	330
0.25	150	0.62	94	281	375
0.00	300	0.72	0	432	432

Apart from some rounding problems the results of the table could essentially be concluded with the linear equation

$$2.23 X_1 + 1.39 X_2 = 600$$

expressing the maximal mixtures of numbers of calls (X_1, X_2) for the two services. This linear quality, though not general, seems to hold rather well when n is not too small. Note that the values 2.23 and 1.39 are between the peak and the average rates of their own services and that they are dependent on the link capacity but not on the mixture.

We shall discuss this almost linear quality more in section 5.

So far we have discussed cell losses in terms of the overall blocking probabilities. Now another question is how much the individual cell loss probabilities differ. An alternative to describe the superposition process as having a Poisson or a Binomial distribution is to suggest it is approximately normally distributed

$$\Sigma X_i \sim N(m, \sigma^2) \quad \text{with} \quad m = \Sigma m_i, \quad \sigma^2 = \Sigma \sigma_i^2$$

Then the conditional average contribution from X_i given $\Sigma X_j = c$ is (c = link capacity)

$$E(X_i | \Sigma X_j = c) = m_i + (c-m) \sigma_i^2 / \sigma^2$$

The interpretation is now that during a typical loss period we have a number of consecutive ms:s with a distribution of the mixture of the different types of bursts such that the average number of offered cells of type i is proportional to the quantity above and the cells from different bursts (calls) are so well mixed in their order of arrival that it is reasonable to assume that the long time average number of lost cells of a certain type, i , is also proportional to this quantity.

As a consequence of that, the individual blocking probability can with $\rho = m/c$ as the load be written

$$B_i = B(\rho + (1-\rho)z_i/z) \quad \text{with } z_i = \sigma_i^2/m_i, \quad z = \sigma^2/m$$

This approximation was suggested independently in [3] and [14]. For the case above with equally many calls of peak rates 11 and 3 Mbps both with average rates 1 Mbps we had $\rho = 0.55$ with $B = 10^{-9}$. Now $z_1 = 10$ and $z = 6$ and thus $B_1/B = 0.55 + 0.45 * 10/6 = 1.3$.

In almost every reasonable example we find that the individual cell loss probability is in the same magnitude as the total one even though one stream is much burstier than the others in the mixture. The only exception is when the load ρ is very small and the streams have very different burstiness. Thus individual blocking probabilities are not a great problem.

We shall now end this section with the discussion we started in section 2 on the importance of the cell scale variance in comparison with this burst scale variance, also given in [2]. Assume that the conditions are such that model I can be used and that the buffer size R is large enough so that 1 ms is the right time unit in the burst scale (otherwise similar arguments can be given).

If the peak rate of a certain type of bursts h is greater than 1 cell/ms e.g. 2, it means that the period is shorter than 1 ms, in the example 0.5 ms, and given a certain number of bursts are on, the number of cells arriving in 1 ms will be the expected number. Thus for $h > 1$ the burst scale variance is sufficient. However if h is less than 1 e.g. 0.2 the period is $1/h$ i.e. 5 in our example. Thus given N bursts are on, the variance of arriving cells in a ms is $Nh(1-h)$, where N itself is a random variable with mean and variance equal to $a = m/h$. Then the total variance is

$$\sigma_0^2 = ah(1-h) + ah^2 - ah = m$$

and not mh which holds when $h > 1$. Even though the choice of time unit 1 ms is not obvious, this shows that for bursty streams with a peak rate above a certain value the burst scale variance alone is sufficient. Compare [13].

4. THE CASE WITH MANY BUFFERS

So far we have in the burst scale regarded the system as more or less a pure loss one since the number of buffers, average burst size ratio R/b has been assumed to be small. Let us for a while assume a greater value of R in the same magnitude as b and try to answer how much link capacity we could save by this fact. We shall here only look at this queueing extension of the previous problem in connection with model I, even though something similar but more complex can be done also for model II. The approximation we shall suggest is explained more carefully in [2] and includes several approximation steps. Here we shall only refer to some of those steps. A useful random variable is Z - the maximum number of cells needing buffers simultaneously during a busy period of a corresponding $M/M/n/\infty$ -system for bursts and a quantity of interest in the approximation is

$$e = E(Z-R|Z \geq R)P(Z \geq R)/EZ$$

Then with Y as the accumulated waiting time in a busy period of an $M/M/1/\infty$ -queueing system with service rate 1 and service factor ρ we have approximately that Z is distributed like bY/n .

We can find the first two moments of Y in [9] and then fit those with a distribution being 0 with probability $1-p$ and being exponentially distributed with mean α with probability p . The result of all these operations is that

$$e \approx e^{-t/\alpha} \quad \text{where } t = Rn/b \text{ and } \alpha = (4\rho+1)(1-\rho)^{-3}$$

After some further steps in the approximation we have

$$B = B_R \approx eB_0$$

where B_0 is the previous cell loss probability and e is as above with $\rho = m/c$ i.e. the load.

Thus with R/b not too small the factor e in B_R can contribute to make B_R small and we can let B_0 be higher e.g. 10^{-7} in some cases which means we can offer a higher load. This however will be to the cost of R buffers and an extra maximum delay for certain cells of R/c . (Of course a simple approximation formula like the one above must be modified to be general in the whole parameter space.)

In an example we have link capacity $c = 150$ Mbps = 375 cell/ms and calls having peak and average rates 2 and 0.8 Mbps and mean burst duration equal to 40 ms. By calculating σ^2 and v with these model II parameters we find that they correspond to the model I values $h = 3$ cells/ms, $b = 72$ cells i.e. $n = c/h = 125$

This implies that with $B = 10^{-9}$, $R \approx 0$ gives a load of the link $\rho = 0.59$ while $R = 150$ gives $\rho = 0.63$ and the max extra delay $R/c = 0.4$ ms. Other cases with the same burst duration but with possibly higher burstiness factor e.g. 5 instead of 2.5 or with the traffic offered to a greater link capacity e.g. $c = 600$ instead of 150 Mbps show that savings allowing around 5% more traffic on the link, given $R/c = 0.4$ ms, are typical. With longer max extra delay R/c acceptable the savings could be even more important. On the other hand with essentially higher values of the mean burst duration the extra savings are more difficult to make. Anyway the possibility of a greater number of buffers is worth considering. In the next sections however we shall assume that R/b is small again.

Even with a lower number of buffers but with a more complex GOS formulation the value of v or d , the mean burst duration, could still be important. With X as the cell loss proportion during a limited time, e.g. an hour, and with demands not only on $EX \leq 10^{-9}$ but also on the corresponding standard deviation, e.g. $DX \leq 10^{-8}$ or $\leq 10^{-7.5}$, this would be the case. This extended GOS formulation is reasonable since cases with 10 lost cells every 10:th hour, 100 every 100:th hour or 1000 lost cells every 1000:th hour all with $EX = 10^{-9}$ still differ a lot in performance and this is taken care of by different DX .

5. THE CALL SCALE PROBLEM

In section 3 we suggested with an example the possibility of finding linear qualities in the burst scale dimensioning. What we actually mean is that the individual virtual resource allocation of a certain type of call should be independent of the mixture of calls on the link with which it shares the capacity. Of course such a quality could never be more than an approximation and its accuracy can not be uniformly good in the whole parameter space. Nevertheless in the very part of this parameter space which is interesting for statistical multiplexing we can, by looking at a number of examples of mixtures, find that this linear quality holds rather well. Similar conclusions are also drawn in e.g. [16]. We can also see that results can rather well be summarized in the following way as in [4].

Given a certain GOS e.g. $B = 10^{-9}$ and that we are in an 'interesting part' of the parameter space the virtual resource allocation for a call with parameters either in terms of (m, σ^2) or (m, h) is approximately

$$k = 1.2m + 60 \sigma^2 / c = 1.2 m(1+50(h-m)/c)$$

$$h/c \leq 0.1 \quad h/m \leq 20 \quad h/k \geq 2$$

The last equality corresponds to the case with on-off bursts and if the expression should exceed 1.2h it is replaced by this. In variable bit rate cases when h but not σ^2 is known we shall have to use the correspondence for the on-off case regarding it as the worst case given (m, h) . The constants 1.2 and 50, which might later be determined with greater accuracy, are of course dependent of B . For e.g. $B = 10^{-5}$ they should rather be 1.1 and 30. However the important thing for the following discussion is that we can determine an expression of the type above which is useful.

Note that even calls with $m = h$ can in mixtures with other types of calls get k -values higher than m and thus also higher than the peak rate h . This is one of the reasons for suggesting that calls suited and not suited for statistical multiplexing should be separated from each other in this type of integration.

Now with these k -values determined we are in this call level problem essentially in the same situation as in a network with circuit switched multi-slot traffic, at least if we could trust the declared parameter values of each arriving call. Thus each virtual k -value calculated from (m, h) corresponds to the bit rate of a circuit switched call. Then we also have to consider if we in the call acceptance principle should include some arrangement to equalize the blocking probabilities for different types of calls. For this problem we refer to e.g. [6].

Anyway here in the call scale several parallel links can of course improve the efficiency of the links given a GOS for call blocking. This quality does not hold in the burst scale. Now, first we shall remember that for the blocking probabilities of the multi-slot case without equalization strategies there exists an exact multi-dimensional product form solution. Thus there is perhaps no need for approximative solutions unless the number of different types of calls is so great that the normalization constant can be inconvenient to treat. Still the following Hayward-like approximation, illustrated here for two but meant to be used for many different types of calls, could give a rough idea of the efficiency of the link also in this call scale sense. Suppose we have two streams and that we know their proportions in call scale individual traffic intensities a_1/a_2 i.e. the mixture and that we can calculate their k -values. Thus we also know the overall k

$$k = (a_1 k_1^2 + a_2 k_2^2) / (a_1 k_1 + a_2 k_2)$$

and the link capacity c (or a multiple of c). Put $n = c/k$ and find the A for which $E(n, A) = 0.01$ e.g. Then A/n is the load and $cA/n = Ak$ is the mean traffic volume we shall distribute among the two streams i.e. $Ak a_1 k_1 / (a_1 k_1 + a_2 k_2)$ to stream 1 which means a traffic intensity $a_1 Ak / (a_1 k_1 + a_2 k_2)$.

6. CALL ACCEPTANCE CONTROL AND CHARGING PRINCIPLES

In the previous section we assumed that at least the two parameters m and h of an arriving call were known. In reality we shall have to face situations where this is not the case. Suppose a call is in some way declaring its m and h . The peak rate h can be policed. Whatever method you use e.g. leaky bucket or moving window, it is possible for the network to notice very soon when the actual peak rate is above the declared one and then reject some of the cells without making any mistake. We could thus in the discussion here regard h as known. The average rate m however is not possible to police in that sense, so we shall have to control it in some other way. For this discussion we shall assume a case with pay load link capacity $c = 100$ Mbps and thus the virtual resource allocation of a call is

$$k = 0.6 m(2+h-m)$$

Now we suggest some different principles. In the first one the call acceptance is based on the declared values of m_i and h_i through k_i and its relation to the total K i.e. whether $k_i \leq c-K$. The m_i -value is for the time being trusted but it is individually measured for each period of a certain length e.g. 10 s and then the charging for each period is based on this measurement. With m_0 as the declared and m as the measured value the charging is suggested to be

$$C = 0.6 m_0(2+h-m_0) \quad m \leq m_0$$

$$C = (2(m-m_0)+m_0)0.6(2+h-m_0) \quad m > m_0$$

When the real value is below the declared one we still have to charge for the declared one since we might have rejected other calls based on that declared value. If the real m is above the declared m_0 the charging must be less favourable than if m_0 would have been correctly declared. The particular choice of the factor 2 (>1) to the $(m-m_0)$ part of the charging might look a bit at random though it has an explanation. For calls with a great value of v , i.e. in terms of on-off calls with (m,h,d) as parameters with long mean burst duration d , it will give measurements as well below as above the declared m_0 even though $m_0 = m$ (the theoretical average rate). This might suggest a special strategy in the declaration of such calls, but with the particular choice of charging principle above (with the factor 2) the optimal declaration is still $m_0 = m$ independently of mean burst duration d (or v). This is shown in the following way. (See also [5]).

Assume that the number of offered cells from a call during a measurement period divided by the length of that period, T , is normally distributed i.e. $X \sim N(m, w^2)$ where $w^2 = v/T$. Then with arbitrary constants a and b and with $u = (m-a)/w$ it generally holds that

$$E(a+b(X-a)^+) = m-wu+bw(u\phi(u)+\varphi(u)) = g(u)$$

Thus the u -value giving a minimal value of $g(u)$ is the one for which $g'(u) = 0$ i.e. $\phi(u) = 1/b$. Then with the particular choice of $b = 2$ we get the optimal $u = 0$ i.e. $a = m$ for any w .

Thus the natural declaration is equal to the optimal one. However the average charging is still higher with a great value of d . With T as the length of the measurement period we have

$$E(C_0) = 0.6m(2+h-m)(1+(4hd/\pi mT)^{0.5}(h-m)/h)$$

Now look at an example showing how unfavourable a wrong declaration of m_0 might be (Let d/T be small). With $h = 2$, $m = 0.5$ we have $k = 1.05$ but with $m_0 = 0.2$ i.e. $k_0 = 0.46$ we will get $C_0 = 1.82$. Thus the wrong m_0 might decrease the risk of being rejected but then the charging is almost doubled. This principle ought to make the majority of calls to be well declared.

Another possibility is to accept a new call only when $1.2 * h$ is less than $c-K$. Then it is no advantage with a too low k_0 in the call acceptance. Otherwise k_0 and C is as before.

In the second call acceptance principle we don't have any declaration of m_0 but only of h . The charging is still based on individual measurement of m for each period and the call acceptance is based on mean and variance measurements of the total traffic stream.

Let x_i be the total number of cells offered to the link in the i :th ms and let the measurement period be e.g. 1 s (or 10 s). Then

$$K = 1.2\bar{x} + (\sum x_i^2 - n\bar{x}^2)60/c(n-1)$$

where in this case $c = 100$ and $n = 1000$, is the estimated virtual resource allocation for the link. The K -value for the latest full second is used and the capacity $c-K$ can be used for new calls which however when arriving will be given the resource allocation $1.2 h$. However in the next second they are included in the measurement of the total K and thus their actual effect on K is used during the major part of those call durations.

The third variant is to assume that no individual measurements are possible. The measurements of the total process and the call acceptance based on the peak rates only could be like in the second principle but the charging can only be based on peak rates and thus the general savings due to some bursty calls will be shared equally among bursty and non bursty calls. This policy might tend to less bursty calls in the network and the importance of statistical multiplexing will decrease.

7. CONCLUSIONS ON THE STATISTICAL MULTIPLEXING IN THE ATM NETWORK

Now is statistical multiplexing worthwhile? First of all I think that one should have two sections in the network, one with and one without statistical multiplexing. Calls which are not particularly bursty or at least have a burstiness factor near 1, but also calls which are bursty but don't have the peak rate, pay load of the link ratio h/c small enough should be separated from the other calls for which statistical multiplexing is worthwhile.

Now using $k = 1.2m(1+50(h-m)/c)$ as above letting the pay load of the link c be 125 and 500 Mbps we can see the difference for some cases

h	2	2	2	2	8	8	8	8
m	.5	.5	.25	.25	2	2	1	1
c	125	500	125	500	125	500	125	500
k	.96	.69	.51	.35	8.16	3.84	4.56	2.04

We can see that in one of these cases the statistical multiplexing is not worthwhile though a burstiness factor of 4. On the other hand it is important to let the savings from all the other cases (and similar ones) help to justify the ATM-network. We should also try to develop ATM towards 600 Mbps links ($c = 500$).

ATM without statistical multiplexing gives hardly any savings, at least not in the middle of the network. Consider a circuit switched network with channels of band widths with a factor 4 between them e.g. 2, 8 and 32 and equal traffic intensities for calls with 2, 8, 32 Mbps as for e.g. 3, 4, 6 and 12, 16, 24 Mbps. Those between the standard band widths are using the one closest above. One call of each type above 2 gives 105 to the link capacity of 160. This loss of 34 % is probably in the same magnitude as the over head losses in ATM-networks. Thus ATM without statistical multiplexing is not such a good idea.

REFERENCES

- [1] Lindberger, K.: 'Traffic Analysis of Statistical Multiplexing in ATM-Networks'. COST224 TD(89)006.
- [2] Lindberger, K.: 'Statistical Multiplexing on ATM-links with a Great Number of Buffers'. COST224 TD(89)055.
- [3] Lindberger, K.: 'Individual Cell Loss Probabilities for the Integrated Substreams on an ATM-link'. COST224 TD(90)002.
- [4] Lindberger, K.: 'The Linear Qualities from the Burst Scale Dimensioning of Mixed Types of Calls in the Call Acceptance and Charging Principles'. COST224 TD(90)003.
- [5] Lindberger, K.: 'Call Acceptance and Charging Principles in Connection with Statistical Multiplexing on ATM-links'. COST224 TD(90)021.
- [6] Lindberger, K.: 'Blocking for Multi-Slot Heterogeneous Traffic Streams Offered to a Trunk Group with Reservation'. 5:th ITC Seminar Lake Como (1987). North Holland 1988 p.151-160.
- [7] Feller, W.: 'An Introduction to Probability Theory and its Applications II'. Wiley New York (1966).
- [8] Smith, W.L.: 'Renewal Theory and its Ramifications'. J Roy. Statist. Soc. B. Vol 20, (1958) 243-302.
- [9] Peiram, L.: 'The Simultaneous Distribution of the Length, the Number of Served Customers and the Accumulated Waiting Time of a Busy Period in an M/M/1-system'. Techn. Report Ericsson Dtm4120 (1974) (in Swedish).
- [10] Eckberg, A.E.: 'The single server queue with periodic arrival process and deterministic service time'. IEEE Trans. Commun. Vol 27 556-562 (1979).
- [11] Roberts, J.W., Virtamo, J.T.: 'The Superposition of Periodic Cell Arrival Processes in an ATM Multiplexor'. IEEE Trans. Commun. (To appear)
- [12] Jacobsen, S.B., Moth, K., Dittmann, L. and Sällberg, K.: 'Load Control in ATM Networks'. ISS90 Stockholm (1990).
- [13] Norros, I., Roberts, J.W., Simonian, A., Virtamo, J.: 'The Superposition of variable bit rate sources in an ATM multiplexor.' Submitted to IEEE J SAC
- [14] Virtamo, J., Norros, I.: 'Who loses Cells in the Case of Burst Scale Congestion?' COST224 TD(90)004. Submitted to ITC 13.
- [15] Simonian, A., Virtamo, J.: 'Transient and Stationary Distributions for Fluid Queues and Input Processes with a Density'. COST224 TD(90)007.
- [16] Appleton, J.: 'Modelling a Connection Acceptance Strategy for ATM Networks'. 7:th ITC Seminar New Jersey (1990).